

```
1 ---
2 title: "Module 9 Lab Activity"
3 subtitle: Gemma Wallace & Neil Yetz
4 output:
5   html_document:
6     df_print: paged
7     toc: yes
8   html_notebook:
9     toc: yes
10 ---
```

```
11
12 # Load libraries
13 ```{r, message=FALSE}
14 library(tidyverse)
15 library(psych)
16 library(olsrr)
17 library(GGally)
18 ```
```

```
19
20 # Import data
21 ```{r,message=FALSE}
22 bac <- read_csv("bac_obs.csv")
23 ```
```

```
24
25 # Describe the data
26 ```{r,message=FALSE}
27 describe(bac)
28 ```
```

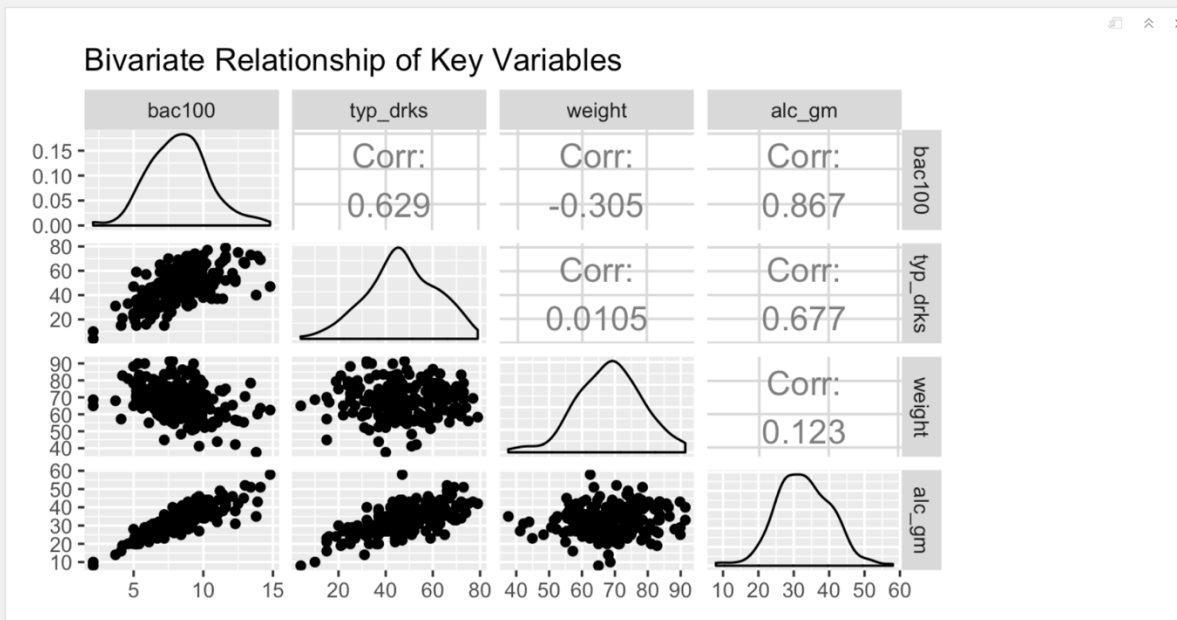
	vars <dbl>	n <dbl>	mean <dbl>	sd <dbl>	median <dbl>	trimmed <dbl>	mad <dbl>	min <dbl>	max <dbl>
id	1	200	100.50	57.88	100.50	100.50	74.13	1.00	200.00
weight	2	200	68.46	9.92	68.90	68.52	9.56	37.60	91.40
typ_drks	3	200	46.86	14.67	46.50	47.11	14.08	4.00	79.00
alcexp	4	200	4.09	0.78	4.13	4.11	0.77	2.01	6.06
pmood	5	200	5.12	1.40	5.00	5.09	1.48	1.00	9.00
absorb	6	200	4.69	0.91	4.64	4.68	0.95	2.67	6.80
alc_gm	7	200	32.79	7.73	33.00	32.76	8.90	8.00	58.00
bac	8	200	0.08	0.02	0.08	0.08	0.02	0.02	0.15

8 rows | 1-10 of 13 columns

```

30 # Mutate BAC variable
31 ```{r}
32 bac <- mutate(bac, bac100 = bac*100)
33 ```
34
35 # Get Scatterplot matrix
36 ```{r, message = FALSE}
37 scatterplot2 <- ggpairs(bac, columns = c("bac100", "typ_drks", "weight", "alc_gm"),
38 upper = list(continuous = wrap("cor", size=5)),
39 title = "Bivariate Relationship of Key Variables", progress = ggmatrix_progress(clear = TRUE))
40
41 print(scatterplot2)
42
43 ```

```



The plot above presents the distributions (via a density plot) for each individual variable, along with scatterplots and correlations for each pair of included variables. All four variables appear to have a roughly normal distribution. The plot indicates that there are large, positive correlations between `alc_gm` and `bac100`, between `typ_drks` and `bac100`, and between `typ_drks` and `alc_gm`.

```

46 # Fit Regressions|
47 ## Regress Bac100 on weight and typ_drks
48 ```{r}
49 m1 <- lm(data = bac, bac100 ~ weight + typ_drks)
50 ols_regress(m1)
51 ```

```

Model Summary							
R	0.702	RMSE	1.548				
R-Squared	0.493	Coef. Var	18.687				
Adj. R-Squared	0.487	MSE	2.397				
Pred R-Squared	0.476	MAE	1.234				
RMSE: Root Mean Square Error							
MSE: Mean Square Error							
MAE: Mean Absolute Error							
ANOVA							
	Sum of Squares	DF	Mean Square	F	Sig.		
Regression	458.540	2	229.270	95.641	0.0000		
Residual	472.248	197	2.397				
Total	930.788	199					
Parameter Estimates							
model	Beta	Std. Error	Std. Beta	t	Sig.	lower	upper
(Intercept)	8.563	0.839		10.212	0.000	6.909	10.216
weight	-0.068	0.011	-0.311	-6.133	0.000	-0.090	-0.046
typ_drks	0.093	0.007	0.632	12.460	0.000	0.078	0.108

52

### 8d

The intercept, 8.563, represents the expected value in y (in this case bac100) when both predictors (weight and typ\_drks) are equal to zero.

The beta estimate for weight, -0.068, is the expected change in bac100 for a one-unit increase in weight *while holding constant* (i.e., controlling for) typ\_drks. In other words, when typ\_drks is controlled for, a one-unit increase in weight is expected to result in a 0.068 unit decrease in bac100. The confidence interval for the weight beta does not include zero and  $p < 0.001$ , indicating that this effect is statistically significant.

The beta estimate for typ\_drks, 0.093, is the expected change in bac100 for a one-unit increase in typ\_drks *while holding constant* weight. In other words, when weight is controlled for, a one-unit increase in typ\_drks is expected to result in a 0.093 increase in bac100. The confidence interval for the typ\_drks beta does not include zero and  $p < 0.001$ , indicating that this effect is statistically significant.

```
54 ## Regress Bac100 on alc_gm and typ_drks
```

```
55 ```{r}
```

```
56 m2 <- lm(data = bac, bac100 ~ alc_gm + typ_drks)
```

```
57 ols_regress(m2)
```

```
58 ```
```

Model Summary							
R	0.869	RMSE	1.077				
R-Squared	0.754	Coef. Var	13.001				
Adj. R-Squared	0.752	MSE	1.160				
Pred R-Squared	0.748	MAE	0.791				
RMSE: Root Mean Square Error							
MSE: Mean Square Error							
MAE: Mean Absolute Error							
ANOVA							
	Sum of Squares	DF	Mean Square	F	Sig.		
Regression	702.202	2	351.101	302.586	0.0000		
Residual	228.586	197	1.160				
Total	930.788	199					
Parameter Estimates							
model	Beta	Std. Error	Std. Beta	t	Sig.	lower	upper
(Intercept)	0.278	0.335		0.830	0.408	-0.382	0.937
alc_gm	0.228	0.013	0.814	16.962	0.000	0.201	0.254
typ_drks	0.012	0.007	0.078	1.634	0.104	-0.002	0.025

59

### 8d

The intercept, 0.278, represents the expected value in bac100 when both alc\_gm and typ\_drks equal zero.

The beta estimate for alc\_gm indicates that, when controlling for typ\_drks, a one-unit increase in alc\_gm is expected to result in a 0.228 increase in bac100. The confidence interval for this estimate does not include zero and  $p < 0.001$ , indicating that this effect is statistically significant.

The beta estimate for typ\_drks indicates that, when controlling for alc\_gm, a one-unit increase in typ\_drks is expected to result in a 0.012 increase in bac100. The confidence interval for this estimate includes zero and  $p > 0.100$ , indicating that this effect is not statistically significant.

```

62 ## Regress BAC100 on weight, alc_gm, and typ_drks
63 ```{r}
64 m3 <- lm(data = bac, bac100 ~ alc_gm + weight + typ_drks)
65 ols_regress(m3)
66 ```

```

Model Summary							
R	0.961	RMSE	0.604				
R-Squared	0.923	Coef. Var	7.290				
Adj. R-Squared	0.922	MSE	0.365				
Pred R-Squared	0.919	MAE	0.447				
RMSE: Root Mean Square Error							
MSE: Mean Square Error							
MAE: Mean Absolute Error							
ANOVA							
	Sum of Squares	DF	Mean Square	F	Sig.		
Regression	859.279	3	286.426	785.073	0.0000		
Residual	71.509	196	0.365				
Total	930.788	199					
Parameter Estimates							
model	Beta	Std. Error	Std. Beta	t	Sig.	lower	upper
(Intercept)	6.056	0.336		18.036	0.000	5.394	6.718
alc_gm	0.253	0.008	0.903	33.142	0.000	0.238	0.268
weight	-0.091	0.004	-0.416	-20.749	0.000	-0.099	-0.082
typ_drks	0.003	0.004	0.022	0.829	0.408	-0.005	0.011

67

### 8d

The intercept, 6.056, represents the expected value of bac100 when all three model predictors (alc\_gm, weight, and typ\_drks) are equal to zero.

The beta estimate for alc\_gm indicates that, while controlling for both weight and typ\_drks, a one-unit increase in alc\_gm is expected to result in a 0.253 increase in bac100. The confidence interval for this estimate does not include zero and  $p < 0.001$ , indicating that this effect is statistically significant.

The beta estimate for weight indicates that, while controlling for both alc\_gm and typ\_drks, a one-unit increase in weight is expected to result in a 0.091 decrease in bac100. The confidence interval for this estimate does not include zero and  $p < 0.001$ , indicating that this effect is statistically significant.

The beta estimate for typ\_drks indicates that, while controlling for both alc\_gm and weight, a one-unit increase in typ\_drks is expected to result in a 0.003 unit increase in bac100. The confidence interval for this estimate includes zero and  $p > 0.100$ , indicating that this effect is not statistically significant.

### 8e – part i

The model  $R^2$  value increased across the three models, indicating that each subsequent model explained more variance in bac100 than the previous model. In general, increasing the number of predictors in your model increases the model  $R^2$ , since more predictors often means that a greater proportion of the unique variance in y is accounted for. In the next module, we'll discuss

how to balance maximizing model explanatory power without overfitting to your data via hierarchical regression.

## 8e – part ii

MLR is essentially a semi-partial correlation, in which the beta estimate for each predictor represents the expected relation between that variable and y when controlling for (or partialing out) the impacts of other predictors. In this example, the beta estimate for typ\_drks is significant when only controlling for the effects of weight (model 1), but it is not significant when controlling for either alc\_gm (model 2) or both weight and alc\_gm (model 3). This indicates that typ\_drks does not explain a significant amount of the unique variance in bac100 when the effects of alc\_gm on typ\_drks is removed.

```

80
81
82 # Subset data for correlation matrix
83 ```{r}
84
85 bac_subset <- select(bac, bac100, alc_gm, weight)
86 cor_matrix <- cor(bac_subset)
87
88
89
90
91
92
93 # Calculate regression from correlation matrix
94 ```{r}
95 set.cor(y = ("bac100"), x = c("alc_gm", "weight"), z = NULL, data = cor_matrix)
96

```

	slope <dbl>	VIF <dbl>
alc_gm	0.92	1.02
weight	-0.42	1.02

2 rows

The slope values for alc\_gm & weight are equivalent to the standardized slopes in our regression model above.

	R <dbl>	R <sup>2</sup> <dbl>	R <sub>uw</sub> <dbl>	R <sub>2uw</sub> <dbl>
bac100	0.96	0.92	0.88	0.78

1 row

The R and R<sup>2</sup> values are equivalent to the multiple regression model with the exact same parameters specified.

This method may be preferable if you need to weight one variable more than the other. MLR uses optimal weights to reduce the squared errors as much as possible. However, if you, as the researcher, know that alc\_gm should be weighted differently, then you would need to use the correlation matrix and perform matrix algebra to weight them appropriately.